

## DOCUMENT RESUME

ED 238 299

FL 014 165

AUTHOR Herzog, Martha; Shaw, Victor  
TITLE An Indirect Test of Oral Proficiency.  
PUB DATE 83  
NOTE 19p.; Paper presented at the Annual Meeting of the Modern Language Association (100th, New York, NY, December 27-30, 1983).  
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)  
  
EDRS PRICE MF01/PC01 Plus Postage.  
DESCRIPTORS Adult Education; Arabic; Audiotape Recordings; Classification; Class Size; Difficulty Level; French; German; Interviews; Korean; \*Language Proficiency; \*Language Tests; \*Oral Language; Polish; \*Questioning Techniques; Rating Scales; Role Playing; Russian; Scoring; \*Second Language Instruction; Serbocroatian; Speech Skills; \*Test Construction; Testing; Test Validity  
  
IDENTIFIERS Defense Language Institute CA

## ABSTRACT

The development of an indirect oral proficiency test for foreign languages by the Defense Language Institute is described. The technique uses taped questions to elicit answers, also taped, that are evaluated according to the Foreign Service Institute's oral proficiency rating scale of 0 to 5. Tape recordings are used because of the large numbers of individuals tested, and questions are carefully phrased to elicit relevant language and compensate when possible for the lack of interviewer-student interaction. The test covers proficiency to Level 3, beyond which point students are tested individually. Eighteen questions are included. Nine (three at each level) are informational, with response times varying with the difficulty of the linguistic task. Six role-playing situations (two at each level) are presented, with reasonable complications built in to elicit as much language as possible. These questions have been found to be valid despite their limitations. The remaining questions are devoted to the use of pictures to produce language at all levels, counter typical avoidance of certain language features, and extend the language sample. The test takes about 45 minutes, and the examinee moves from Level 1 to 3 through the three alternating formats. It has been developed for use in Russian, Korean, Modern Standard Arabic, Polish, French, German, and Serbo-Croatian, and validated in Russian and Korean. (MSE)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

U.S. DEPARTMENT OF EDUCATION  
NATIONAL INSTITUTE OF EDUCATION  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

↑ This document has been reproduced as  
received from the person or organization  
originating it.

□ Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official NIE  
position or policy.

## AN INDIRECT TEST OF ORAL PROFICIENCY

BY MARTHA HERZOG AND VICTOR SHAW  
DEFENSE LANGUAGE INSTITUTE  
FOREIGN LANGUAGE CENTER

Herzog/Shaw  
TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

ED238299

For about thirty years the federal government has used an eleven point scale based on written definitions as a measure for testing oral proficiency. That scale covers Levels 0 through 5, including plus points, such as 0+, 1+, and so forth.

(Figure 1)

Associated with that scale is a formal test, commonly known as the FSI interview because the Foreign Service Institute developed the fundamental techniques. However, the test is now used by several other government agencies, which call it the Interagency Oral Proficiency Interview.

FL014165

The new name comes from the Interagency Language Roundtable, an organization of all government agencies engaged in foreign language teaching. Members of the Roundtable have refined the original FSI interview, added to the supporting documents clarifying the system, and engaged in research related to oral proficiency.

Furthermore, since 1981, the American Council on the Teaching of Foreign Languages (ACTFL) has been working with the federal government to adapt the eleven point scale to academic needs and to provide training in oral proficiency testing. This work has been widely publicized, and it was a major topic at the annual ACTFL convention in New York in 1982.

(Figure 2)

Our agency, the Defense Language Institute Foreign Language Center (DLIFLC), is a part of the Roundtable, and we have assisted with the ACTFL projects. Since January 1981, DLIFLC has trained and certified approximately 120 members of our own faculty and staff to administer the oral interview in 25 languages; we have also trained about 30 ACTFL representatives in four languages. DLIFLC is intensely interested in the interview as a method of measuring oral proficiency.

Nevertheless, because of the size of our test population, we have also found it necessary to develop an indirect method of testing oral proficiency that makes maximum use of the live interview's elicitation techniques and scoring system.

It would be ideal if we could test every student in a face-to-face interview, not only at graduation, but at significant phases of the course. However, five of the thirty languages that we regularly teach have over four hundred students per year; one language has over a thousand. It is not practical to interview all of these students. In addition the Army requires that graduates be retested every two years, regardless of where they may be stationed at the time. Finally, we must test United States military personnel who acquired a foreign language by other means, such as through formal education, by living abroad, or by growing up in a bilingual family. These people may be located anywhere in the world. The American military needs an indirect test of oral proficiency, despite the obvious limitations.

The list of limitations begins with the fact that any taped test inevitably produces an impersonal atmosphere. The live interview always starts with a warm-up period to put the examinee at ease, to get him used to speaking in the target language, and to create the proper psychological setting before there is any serious questioning. That friendliness is missing when one speaks into a tape recorder.

Even more of a disadvantage, from the point of view of the tester, is the lack of opportunity to ask follow-up questions on tape and to build upon the speech sample that the examinee produces in order to make him explain, rephrase, support opinions, and so forth. Interaction is an integral part of the live interview, while it is completely missing from a taped test. Therefore, the cautious examinee who monitors every word and avoids grammatical constructions that he knows he cannot control will exercise this caution far more successfully with a machine than with a team of experienced testers.

However, a great deal of thought went into the design of the indirect speaking test in an effort to overcome some of the limitations. The test designers and the writers of the target language material are all certified testers in the Interagency Oral Proficiency Interview system. They have administered hundreds of interviews and scored even more hundreds of tapes that use various types of questions.<sup>1</sup>

We decided at the outset not to try to test beyond the government's Level 3 with an indirect test. Any examinee who scores a Level 3 on the taped test will be eligible for a complete interview either in person or by telephone, depending on the location.

(Figure 3)

Level 3 is actually very high. A Level 3 can discuss both concrete and abstract subjects in either a formal or informal conversational style. After 47 weeks of intensive training, which (excluding homework) takes 30 hours per week for a total of approximately 1,400 hours, most of our graduates remain well below Level 3 in speaking. So do 90% of all military personnel who use a foreign language in their work.<sup>2</sup> Therefore, we can afford to administer interviews to the comparatively small number who do attain this level of proficiency on the taped test.

(Figure 4)

The indirect test contains eighteen questions. There are basically three types of questions; however, the difficulty for each type ranges from Level 1 through Level 3. Scores can be assigned for the plus levels, including 0+, when performance exceeds the base level.

Nine questions--three at each level--are informational in nature. Examinees are asked in the target language on tape to state facts, give explanations, describe, hypothesize. There are progressively more difficult linguistic tasks, directly related to the level definitions. Depending on the difficulty of the task, response time ranges from 25 seconds to 2 minutes; the examinee is told exactly how much time he will be given, and he is urged to respond as fully as possible.

Because there is no opportunity to probe or follow up a short initial response, each question must be carefully written to elicit a speech sample that allows the examinee to use as much relevant language as he can. The Level 3 questions are sufficiently complex that we voice each one twice on the tape. This provides the examinee a moment to think before he responds just as he might do with a more abstract topic in an interview or, for that matter, in a real conversation.

Distributed at the appropriate levels are six role-playing situations--two at each level. Role-playing is an essential part of the interview. It permits the examinee to show how well he can use the target language to solve problems or to get himself out of unexpected situations. In the interview, of course, role-playing relies upon interaction. A simple, straightforward role-playing scenario, such as renting a hotel room, can quickly be

made more complicated if the interview team discovers that the original task was too easy for an examinee. Role-playing can be continuously adjusted so that unanticipated problems confront the interviewee just as they do in everyday life. But because of the unpredictability of examinee response, we cannot interject such complications onto the tape.

However, if we work them out carefully, taped role-plays can offer the examinee the chance to use his imagination, to pretend to be a different person or, perhaps, just to find his way out of an unusual circumstance. The essential requirement is that reasonable complications have to be built into the initial role-playing situation. With that in mind we have been able to create situations in which a person might plausibly speak at some length without getting a rejoinder. Furthermore, if we write the mindset and the rhetorical stance of the speaker into the role-play, we can both control the difficulty level and imply valid reasons for speaking at length. In addition, the explicit taped instructions let the examinee know that he has to bring as much language as possible into the role-play.

In spite of their imperfections, taped role-plays do work. In our past research, taped role-playing situations showed a strong correlation with the global interview score.

The basic role-playing scenario is written in English in the test booklet and also voiced in English on the tape. This is similar to what happens during an interview; there the situation is described in English (often typed on an index card), and the examinee is given a few seconds to prepare for his task.

Like the informational questions, the role-plays are sequenced according to difficulty, moving gradually from Level 1 to Level 3. Level 1 essentially requires the transmittal of information in a survival situation. Level 2 requires concrete language, while Level 3 demands that the speaker give reasons for his actions or preferences.

The third part of the test format involves the use of several sets of pictures drawn in the booklet. The pictures enable us to counter typical avoidance strategies. Experienced target language interviewers have observed that large numbers of examinees tend to avoid certain language features during interviews. We can probe for this particular language through the use of pictures that must be explained. Like the other parts of the test, the pictures present increasingly more difficult linguistic tasks. The response time allowed ranges from one to two minutes, depending on the level.

An advantage of pictures is that they enable the lowest level speaker--the Level 1 or even the 0+--to produce some language. He may not perform the expected task; however, at the 0+ level, for example, he can enumerate the things that he sees--boy, dog, car. The Level 2 speaker can describe the persons, objects, and actions through connected discourse. The Level 3 can say precisely what is depicted, perhaps explaining why it is unusual, if that is the case, or hypothesizing about his own behavior in similar circumstances.

Sets of pictures provide a versatile measure for speaking and, combined with the other two formats, yield a good sample of language that can be rated according to the same scale used for the interview.

The taped oral proficiency test is very thorough. The examinee is in the language lab for about forty-five minutes. He moves gradually from Level 1 to Level 3 through the three alternating formats. He has the opportunity to demonstrate what he can do with the language on a variety of topics. And he must perform a wide range of tasks. All these tasks are sensitive to the particular problems adult Americans have when learning the target language being tested.

Our goal is to have each examinee record what is called a "ratable sample". In an interview, of course, the testers can continue to ask questions and present tasks until they are confident the examinee has demonstrated his typical functional ability with the target language as well as the best he can do and clear-cut examples of some things he cannot do.

The indirect test attempts to do the same thing. By gradually moving from the easier to the more difficult tasks, we offer a chance for success. Because the test measures a level higher than that achieved by most of our graduates, even at the end of a lengthy intensive course, there is also a probability of "linguistic breakdown", as it is called in analyses of live interviews. The Level 1 or Level 2 speaker is encouraged to try to deal with the subject matter of the Level 3 questions; but he will probably not control the grammar or the vocabulary needed to complete these tasks successfully.

The tests were developed by experienced target language interviewers and are also scored by interviewers, using exactly the same criteria as used for scoring a direct test. After two trained interviewers listen to an entire tape, they assign a global rating. Questions are not scored separately or weighted; nor are there separate scores for pronunciation, grammar, and other factors. Instead, the interviewers make use of their extensive training and experience in relating such factors as vocabulary, grammar, fluency, etc. to a global rating.

Highly experienced interviewers can score a taped test after one hearing; less experienced testers might need to listen a second time, particularly to problem cases. If voice-activated equipment is used for recording examinee tapes, scoring time can be saved. But, even when older lab equipment is used, the taped test has certain advantages.

Examinees can be tested at any time at any location that has any form of recording equipment. A target language speaker need not be present to administer the test. Tapes can be scored at the convenience of the interviewers. Also, in a military environment there is a great deal of respect for standardization. While the test presently consists of two equivalent alternate forms, every examinee gets the same chance to answer the same number of questions at the same level of difficulty. The charge of subjectivity that is often leveled at even the best speaking tests, such as the interview, has less validity for the taped test. The indirect test does not solve every problem, but it solves some major ones.

So far the Defense Language Institute has developed taped speaking tests in Russian, Korean, Modern Standard Arabic, Polish, French, German, and Serbo-Croatian. The Russian test served as the prototype, both in design and developmental procedures. It was produced in two alternate forms in 1982 and validated with a population of 118 examinees in the United States and Europe in 1983. About half of the population took Form A of the taped test and half took Form B. All examinees were administered the Interagency Oral Proficiency Interview as an external criterion. The Korean test was validated with 142 examinees in the United States and Korea in 1983. The validation statistics are shown in Figures 5 and 6. The high correlations of each of these tests with the interview indicate that DLIFLC has developed a good method for testing oral proficiency indirectly.

(Figure 5)

(Figure 6)

Footnotes

<sup>1</sup>Substantial work has been done on question types by Pardee Lowe, Jr. Useful material can be found in his Handbook on Question Types and Their Use in LS Oral Proficiency Tests. Washington, D.C.: CIA Language School. Further information is available in his article, written with Ray T. Clifford, entitled "Developing an Indirect Measure of Overall Oral Proficiency" published in Frith, James R., ed. 1980. Measuring Spoken Language Proficiency. (Papers from a preconference workshop sponsored by Georgetown University Language Roundtable on Languages and Linguistics and U.S. Government Interagency Language Round Table.) Washington, D.C.: Georgetown University.

<sup>2</sup>Several years ago John B. Carroll assessed the proficiency of undergraduates majoring in French, German, Russian, and Spanish. After four years of college study, most were rated at 2 or 2+ on the government scale. Carroll's study was reported in "Foreign Language Proficiency Levels Attained by Language Majors Near Graduation from College." Foreign Language Annals I. No. 2 (1967): 131-51.

# BASIC SKILL LEVELS

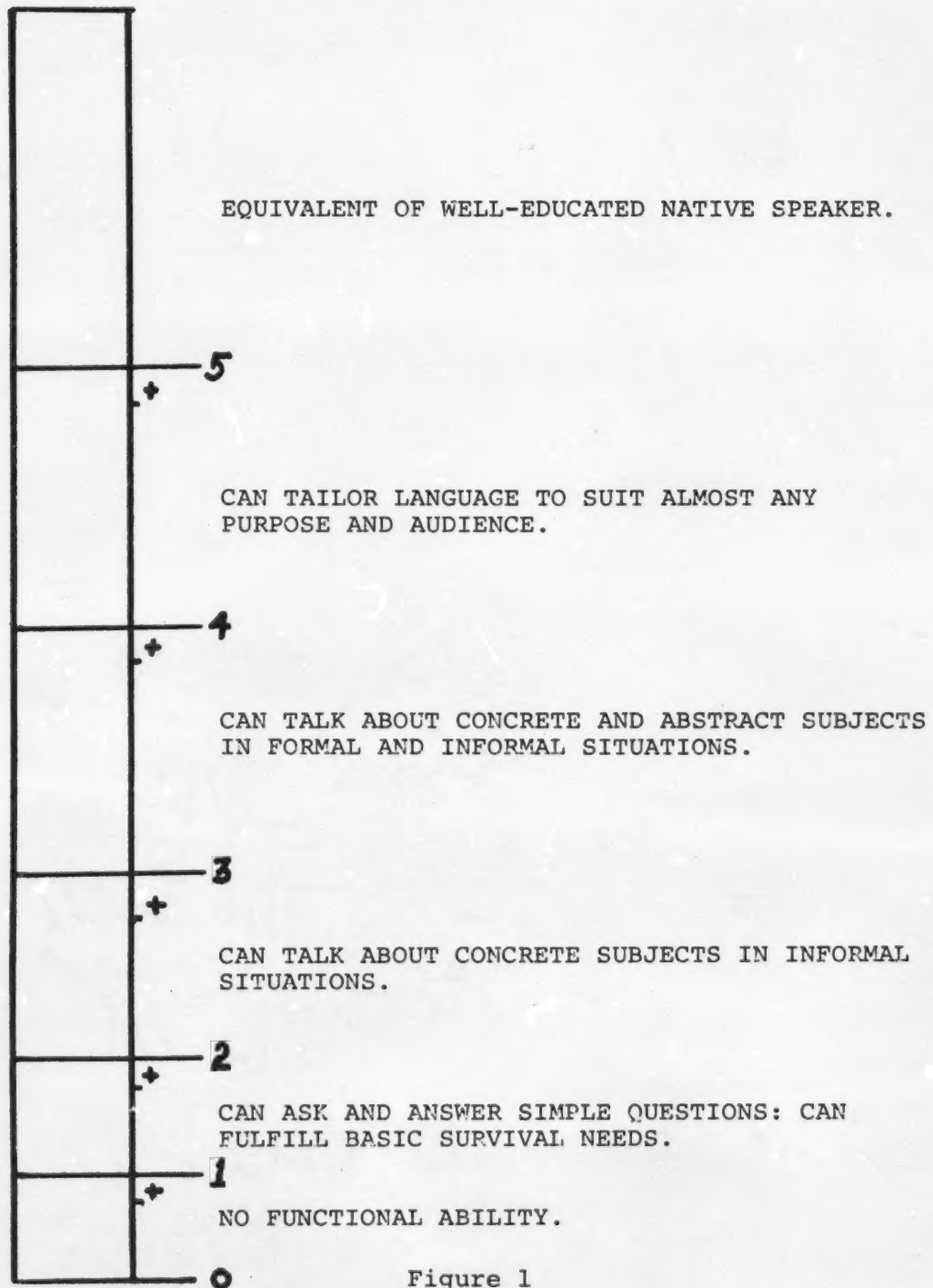


Figure 1

| <u>ACTFL SCALE</u>  | <u>GOVERNMENT SCALE</u> |
|---------------------|-------------------------|
| SUPERIOR            | 3 AND ABOVE             |
| ADVANCED PLUS       | 2+                      |
| ADVANCED            | 2                       |
| INTERMEDIATE - HIGH | 1+                      |
| INTERMEDIATE - MID  | 1 (HIGH)                |
| INTERMEDIATE - LOW  | 1 (LOW)                 |
| NOVICE - HIGH       | 0+                      |
| NOVICE - MID        | 0 (HIGH)                |
| NOVICE - LOW        | 0 (LOW)                 |

Figure 2

### THE LEVEL 3 SPEAKER

#### FUNCTION

CAN CONVERSE IN FORMAL AND INFORMAL SITUATIONS;  
RESOLVE PROBLEM SITUATIONS;  
DEAL WITH UNFAMILIAR TOPICS;  
PROVIDE EXPLANATIONS;  
DESCRIBE IN DETAIL;  
OFFER SUPPORTED OPINIONS;  
HYPOTHESIZE.

#### CONTENT

PRACTICAL, SOCIAL, PROFESSIONAL TOPICS;  
PARTICULAR INTERESTS;  
SPECIAL FIELDS OF COMPETENCE.

#### ACCURACY

ERRORS NEVER INTERFERE WITH UNDERSTANDING AND RARELY  
DISTURB THE NATIVE SPEAKER. ONLY SPORADIC ERRORS  
IN BASIC STRUCTURES.

INDIRECT ORAL PROFICIENCY  
TEST FORMAT

PART ONE

3 INFORMATION QUESTIONS  
2 ROLE-PLAYING SITUATIONS  
1 SET OF PICTURES TO DESCRIBE

LEVEL 1 (SURVIVAL LANGUAGE)

25 SECOND PAUSE  
1 MINUTE PAUSE  
1 MINUTE PAUSE

PART TWO

3 INFORMATION/EXPLANATION QUESTIONS  
2 ROLE-PLAYING SITUATIONS  
1 SET OF PICTURES TO DESCRIBE

LEVEL 2 (CONCRETE LANGUAGE)

1 MINUTE PAUSE  
90 SECOND PAUSE  
90 SECOND PAUSE

PART THREE

3 ANALYTICAL QUESTIONS  
2 ROLE-PLAYING SITUATIONS  
2 SET OF PICTURES TO DESCRIBE

LEVEL 3 (ABSTRACT LANGUAGE)

2 MINUTE PAUSE  
2 MINUTE PAUSE  
2 MINUTE PAUSE

Figure 4

RUSSIAN  
SPEAKING TEST RESULTS

|                               | INTERVIEW | FORM<br>A | FORM<br>B |
|-------------------------------|-----------|-----------|-----------|
| N                             | 113       | 42        | 57        |
| MEAN                          | 1.60      | 1.63      | 1.55      |
| S.D.                          | .86       | .68       | .86       |
| CORRELATION WITH<br>INTERVIEW |           | .84       | .84       |

Figure 5

KOREAN  
SPEAKING TEST RESULTS

|                               | INTERVIEW | FORM<br>A | FORM<br>B |
|-------------------------------|-----------|-----------|-----------|
| N                             | 142       | 71        | 48        |
| MEAN                          | 1.81      | 1.81      | 1.62      |
| S.D.                          | .61       | .61       | .56       |
| CORRELATION WITH<br>INTERVIEW |           | .90       | .94       |

Figure 6